Hermes: A Multi-Scale Spatial-Temporal Hypergraph Network for Stock Time Series Forecasting

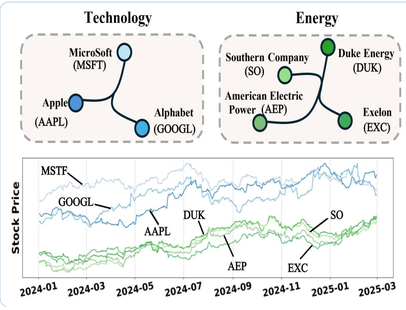
Xiangfei Qiu[1], Liu Yang[1], Xiangyu Xu[1], Hanyin Cheng[1], Xingjian Wu[1], Rongjia Wu[2], Zhigang Zhang[2], Ding Tu[2], Chenjuan Guo[1], Bin Yang[1], Christian S. Jensen[3], Jilin Hu[1]*

[1] East China Normal University [2] CFETS Information Technology (Shanghai) Co., Ltd [3] Aalborg University * Corresponding author

NeurIPS 2026 WeChat Paper

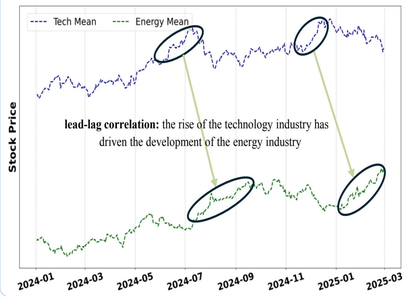
Introduction

Stocks in one industry often move together. Can their movements help forecast returns in another industry days or weeks later?



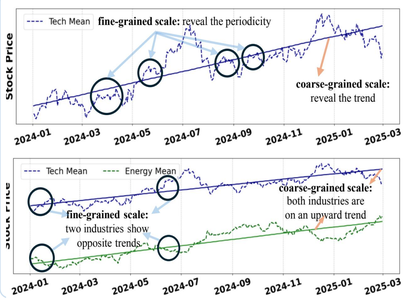
Within an industry

An industry hyperedge pools signals from stocks with shared exposure to AI investment.



Across industries

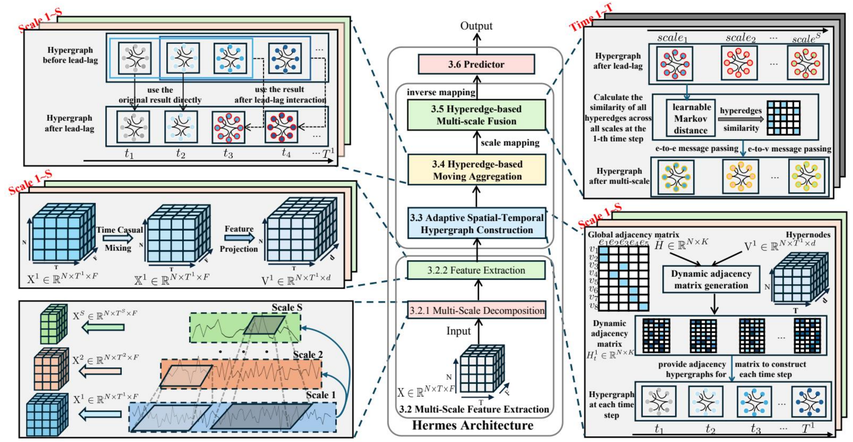
A rise in technology stocks can precede a rise in energy stocks as power demand grows.



Across time scales

A short window captures local swings. A longer window reveals the broader trend.

Hermes architecture



How signals move between industries

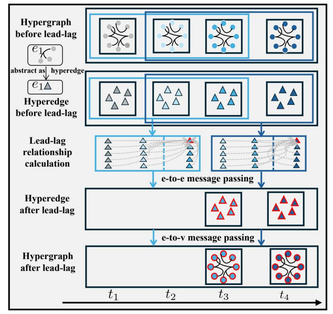


Fig. 1. A three-step window.

Lead–lag aggregation

In a three-step window, each industry compares its current state with earlier industry signals and weights their contributions.

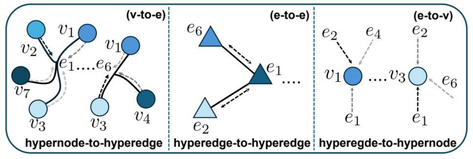


Fig. 2. Stock → industry → industry → stock.

Cross-scale fusion and prediction

Industry hyperedges exchange messages across time scales. Stock-level predictors combine the updated features with the initial features and sum the forecasts across scales.

Experiments

Forecast accuracy across three markets

Models		HGNN				Transformer			MLP			GNN			RNN	
		Hermes	EST.	STHAN.	ASHyper	iTrans.	PatchTST	DUET	SMixer.	DLin.	TMixer.	RSR-I	GAT	RGCN	ALSTM	LSTM
NASDAQ	IC	**0.044**	0.037	0.039	0.041	0.038	0.039	0.039	0.043	0.031	0.041	0.038	0.035	0.034	0.035	0.032
	RIC	**0.538**	0.444	0.451	0.508	0.420	0.496	0.494	0.501	0.453	0.532	0.398	0.377	0.382	0.371	0.354
	SR	**2.161**	1.307	1.416	1.157	1.877	1.261	1.751	1.465	1.778	1.177	1.238	1.233	1.054	0.941	0.892
NYSE	IC	**0.032**	0.030	0.029	0.025	0.023	0.029	0.024	0.029	0.026	0.031	0.026	0.025	0.025	0.023	0.024
	RIC	**0.466**	0.327	0.344	0.250	0.303	0.284	0.327	0.351	0.277	0.314	0.284	0.297	0.275	0.276	0.256
	SR	**1.655**	1.115	1.228	1.208	1.520	0.898	1.361	1.454	1.526	1.556	1.098	1.070	0.932	0.764	0.857
S&P500	IC	**0.050**	0.035	0.037	0.046	0.048	0.046	0.049	0.041	0.043	0.040	0.033	0.034	0.028	0.029	0.031
	RIC	**0.334**	0.241	0.227	0.268	0.242	0.249	0.275	0.262	0.214	0.253	0.200	0.191	0.175	0.181	0.186
	SR	**2.247**	1.547	1.533	2.148	1.326	1.402	1.982	1.586	1.557	2.170	1.437	1.484	1.359	1.298	1.332

Model abbreviations: EST. (ESTIMATE), STHAN. (STHAN-SR), iTrans. (iTransformer), SMX (StockMixer), DLin. (DLinear), and TMX (TimeMixer).

Table 1. Hermes leads 14 baselines on all nine dataset–metric pairs. On NASDAQ, its Sharpe ratio reaches 2.161 versus 1.877 for the strongest baseline, iTransformer.

Ablation study

Ablation	NASDAQ			NYSE			S&P500		
Model Component	IC	ICIR	SR	IC	ICIR	SR	IC	ICIR	SR
w/o Multi-Scale Fusion	0.038	0.464	1.374	0.028	0.421	1.354	0.031	0.248	1.517
w/o Total Multi-Scale	0.030	0.422	0.975	0.023	0.347	1.199	0.032	0.213	1.075
w/o Lead-Lag	0.039	0.472	1.451	0.028	0.420	1.225	0.037	0.201	1.628
w/o Skip-Connection	0.041	0.441	1.655	0.029	0.345	1.449	0.043	0.252	1.653
Only Initial Node Embedding	0.019	0.174	0.695	0.016	0.113	0.449	0.018	0.122	0.659
Hermes (Ours)	**0.044**	**0.538**	**2.161**	**0.032**	**0.466**	**1.655**	**0.050**	**0.334**	**2.247**

Table 2. Removing lead–lag aggregation lowers NYSE Sharpe from 1.655 to 1.225. Removing multi-scale fusion lowers it to 1.354.

Runtime and memory

Dataset	NASDAQ				NYSE				S&P500			
Property	STHAN.	EST.	ASHyper	Hermes	STHAN.	EST.	ASHyper	Hermes	STHAN.	EST.	ASHyper	Hermes
Training Time	257 ms	1,267 ms	288 ms	215 ms	384 ms	1,462 ms	378 ms	282 ms	297 ms	1,317 ms	314 ms	211 ms
Inference Time	52 ms	163 ms	58 ms	42 ms	73 ms	241 ms	71 ms	61 ms	32 ms	109 ms	36 ms	25 ms
Memory Usage	2.15 GB	9.62 GB	1.99 GB	0.98 GB	2.82 GB	10.25 GB	2.17 GB	1.14 GB	1.23 GB	4.52 GB	1.51 GB	0.54 GB

Table 3. At batch size 1, Hermes runs S&P 500 inference in 25 ms and uses 0.54 GB of memory.